

УДК 004.91

О. В. Мороз

ТЕЗАУРУС ЛІНГВІСТИЧНИХ ТЕРМІНІВ: МЕТОДИКА СТВОРЕННЯ І СТРУКТУРА

The paper describes the method of construction and structure of linguistic thesaurus. The main criterion for the selection of terms to the thesaurus was an information criterion. Building a thesaurus provides disclosure of the basic types of relations between concepts; the main ones are correlation, synonymy, hyperonymy-hyponymy (genus-species) and holonymy-meronymy (part-whole). The entry consists of nine items and is formalized encyclopedic description of the term. The thesaurus model and method of its construction can be considered in two aspects: as a linguistic description of the thesaurus as an instrument of a information retrieval and as an informational-logical description of linguistic terminology.

Keywords: *thesaurus, semantic relations, dictionary entry, formal description, terminology.*

Описано методику побудови і структуру тезауруса лінгвістичних термінів. Головним критерієм відбору термінів до тезауруса був критерій інформаційний. Побудова тезауруса передбачає розкриття основних типів відношень між поняттями, основними з яких є кореляція, синонімія, гіпонімія (рід-вид) і голонімія-меронімія (частина-ціле). Словникова стаття складається з дев'яти пунктів і є формалізованим енциклопедичним описом терміна. Запропоновану модель тезауруса і методику його побудови можна розглядати у двох аспектах: як лінгвістичний опис інформаційного пошукового тезауруса і як інформаційно-логічний опис лінгвістичної термінології.

Ключові слова: *тезаурус, семантичні відношення, словникова стаття, формалізований опис, термінологія.*

Побудова онтологій предметних областей починається із побудови тезауруса цієї предметної області. Серед усіх наукових галузей, для яких будуються онтології, галузь прикладної лінгвістики найменше досліджена. Для створення алгоритму побудови такого тезауруса лінгвістичних термінів і вибору відповідних засобів необхідно усвідомити, що основою для тезаурусного опису термінології, зокрема термінології лінгвістичної науки, є методика лінгвістичного конструювання. Лінгвістичне конструювання є "сукупністю узагальнених способів і прийомів комбінування "зразків вирішення проблем", екстраполяції існуючих теоретичних і практичних результатів, отриманих у різних галузях лінгвістики, та їх прямого чи евристичного використання для подолання труднощів, які виникають... при побудові нових лінгвістичних об'єктів" [1].

Вивчаючи певний науковий напрям, дослідникові необхідно опанувати термінологію цього напрямку, оскільки поняття, виражені термінами і семантичні зв'язки між ними, становлять теоретичну базу будь-якої науки. Водночас вивчення тої чи іншої предметної області передбачає вміння описувати її комплексно. Такий комплексний підхід до вивчення предметної області дає дослідникові змогу сформулювати чітке уявлення про систему понять цієї області. Дослідникові необхідно зібрати дані: по-перше, про всі об'єкти, процеси і явища, істотні для досліджуваної галузі, по-друге, про відношення між цими об'єктами чи їх частинами, і по-третє, про реальні і потенційні взаємозв'язки між об'єктами, які виникли у процесі діяльності дослідника. Класичний процес вивчення предметної області передбачає поступове засвоєння термінології, а зв'язки між поняттями, виражені термінами і оточення, у якому вони трапляються, стають зрозумілими лише після закінчення курсу. Натомість тезаурусний метод, який останнім часом стали регулярно застосовувати для опису найрізноманітніших предметних галузей, таких як

© О. В. Мороз, 2013

методика вивчення іноземної мови, корпусна лінгвістика, екологія, освітнє право, морфологія англійської мови та багато інших, уможливило комплексний і об'єктивний опис системи понять тієї чи іншої галузі науки, а після опису – і моделювання цієї системи. Такий опис детально розглянуто у роботах [2–6].

Зазначимо, що тлумачення того чи іншого терміна часто передбачає різні його визначення. Ці відмінності можна пояснити існуванням різних лінгвістичних шкіл та груп. Однак відмінність у підходах до тлумачення тих чи інших мовних явищ аж ніяк не заперечує таких дефініцій, які ґрунтуються на істотних ознаках досліджуваних явищ. Ми брали до уваги основні властивості термінів: системність, усталеність і регулярність зв'язків, відсутність експресивності, об'єктивність опису поняття [7].

У нашому дослідженні розроблений тезаурус лінгвістичних термінів виконує подвійну функцію: здійснює опис і систематизацію лінгвістичної термінології, а також використовується у семантичній метриці для пошуку релевантної інформації. У цій статті розглянуто першу функцію тезауруса.

Постановка задачі. Розробити тезаурус предметної галузі на основі “терміносистеми”, проаналізувавши корпус наукових, навчальних та довідкових джерел, у яких зафіксовані терміни досліджуваної предметної галузі.

Викладення основного матеріалу. Для побудови тезауруса предметної галузі (у нашій роботі це – лінгвістична наука) необхідно сформулювати перелік словникових одиниць і обмежити лексичний матеріал, який стане основою для словника тезаурусного типу. Необхідно окреслити “терміносистему” предметної галузі, проаналізувавши корпус наукових, навчальних та довідкових джерел, у яких зафіксовані терміни досліджуваної предметної галузі, у нашому випадку це теоретичне та прикладне мовознавство.

Матеріалом для побудови тезауруса лінгвістичної термінології стали ключові слова до анотацій статей, вміщених у трьох фахових виданнях ВАК України з мовознавства у період 2009–2011 рр. Цими виданнями є “Вісник Волинського національного університету ім. Лесі Українки” (тепер Східноєвропейський університет ім. Лесі Українки), збірники наукових праць “Нова Філологія” і “Лінгвістика”, які видають Запорізький національний університет і Луганський національний університет ім. Тараса Шевченка відповідно. Загальна кількість опрацьованих анотацій – 2032.

Наступним етапом стало формування алфавітного переліку ключових слів, вилучених із анотацій публікацій, розміщених у вищезазначених наукових виданнях. Далі з цього переліку необхідно було відібрати слова – кандидати в терміни тезауруса. Для цього ми проаналізували низку українських та зарубіжних паперових та електронних довідників та словників лінгвістичної термінології. Серед них варто зазначити такі: “Сучасна лінгвістика: термінологічна енциклопедія” О. О. Селіванової (2006), “Словник термінів міжкультурної комунікації” Ф. С. Бачевича (2007), “Краткий словарь когнитивных терминов” О. Кубрякової (1997), “Большой лингвистический словарь” В. Д. Стариченка (2008), “Экспериментальный толковый словарь стилистических терминов” С. Е. Нікітіної та Н. В. Васильєвої (1996), “Термины и понятия лингвистики: Лексика. Лексикология. Фразеология. Лексикография: Словарь-справочник” Т. В. Жеребило (2011), “Linguistic Thesaurus” Д. Соргеля (2005), онлайн-тезаурус англійської мови Вебстер (<http://www.merriam-webster.com/thesaurus/linguistic>). Головним критерієм відбору був критерій інформаційний, тобто потенційна можливість терміна-кандидата “називати самостійне мовне явище і бути основою для інформаційного запиту, тобто дескрипторздатність терміна” [2]. Тому основну частину відібраних термінів становлять іменники або іменникові словосполучення. Під час відбору термінів ми брали до уваги також факт існування їх визначень у термінологічних слов-

никах чи у статтях, якщо термін-кандидат не був зафіксований у довідковій літературі, а також понятійні зв'язки термінів між собою.

Окрім зовнішніх критеріїв (зафіксованість у термінологічних лінгвістичних словниках, наявність у текстах статей чітких визначень), ми брали до уваги і лінгвістичні критерії (відтворюваність, семантична неподільність словосполучення). Після процедури відбору термінів-кандидатів загальна їх кількість становила 1780 одиниць.

Подальшим кроком стало формування рубрик тезауруса і наповнення їх термінами. Зазначимо, що під поняттям “рубрика тезауруса” ми розглядаємо окремий розділ мовознавства (теоретичного чи прикладного), який, відповідно, має власну термінологію. Формуючи рубрики тезауруса, ми використали класифікацію основних розділів мовознавства, розроблену професором О. О. Селівановою. Згідно з цією класифікацією існують такі розділи сучасного мовознавства: “Лінгвометодологія”, “Семасіологія”, “Лінгвосоціотика й ономасіологія”, “Психолінгвістика”, “Етнолінгвістика та культурологія”, “Соціолінгвістика”, “Когнітивна лінгвістика”, “Граматики”, “Лінгвістика тексту”, “Комунікативна лінгвістика”, “Прикладна лінгвістика” [8].

Більшість названих тут розділів мовознавства подані окремими рубриками у проаналізованій нами науковій періодиці. Однак у зазначених збірниках наукових праць є і рубрики, які репрезентують важливі розділи мовознавства, однак не увійшли до цієї класифікації. Тому, готуючись до укладання тезауруса і формування його структури, ми доповнили класифікацію О. О. Селіванової і запропонували таку структуру тезауруса, яка відображає існуючий поділ лінгвістики на прикладну та теоретичну і водночас охоплює усі рубрики, подані у проаналізованих нами збірниках наукових праць. Отож до розділу “Прикладна лінгвістика” увійшли 7 рубрик: “Корпусна і комп'ютерна лінгвістика”, “Лексикографія”, “Термінознавство”, “Методика”, “Переклад”, “Психолінгвістика”, “Соціолінгвістика”. Розділ “Теоретична лінгвістика” склали 8 рубрик і відповідні їм підрубрики: “Комунікативна лінгвістика” (підрубрики “Міжкультурна комунікація”, “Функціональна лінгвістика. Прагмалінгвістика”, “Дискурсологія”); “Лінгвістика тексту”; “Граматики”; “Когнітивна лінгвістика”; “Етнолінгвістика. Культурологія”; “Лінгвосоціотика”; “Фразеологія. Лексикологія. Діалектологія (однойменні підрубрики і “Семантика”; “Лінгвометодологія” (підрубрика “Теорія мови”).

Наступним етапом нашого дослідження було укладання тезауруса лінгвістичних термінів на основі зібраного матеріалу. Тезаурусний словник є переліком логіко-семантичних відношень між лінгвістичними термінами. Цей тезаурус охоплює не тільки множину окремих термінів, поданих у вигляді алфавітного списку з їхніми тлумаченнями, а й моделі показу зв'язків між термінами. До тезауруса увійшли терміни, переважно іменники або іменникові словосполучення, з прикладної та теоретичної лінгвістики. Побудова тезауруса передбачає розкриття основних типів відношень між поняттями, основними з яких є кореляція, синонімія, гіпонімія (рід-вид) і супідрядність на одному рівні – парціація (частина-ціле). Зміст відношень розширено настільки, щоб можна було охопити максимально широкий пласт термінів, з якими пов'язаний аналізований термін як реєстровий. Нижче подаємо опис словникової статті тезауруса.

Словникова стаття складається з дев'яти пунктів і є формалізованим енциклопедичним описом терміна. Структуру словникової статті в узагальненому виді ми подамо нижче, а тут обмежимося переліком основних пунктів словникової статті і прикладами.

1. Назва розділу лінгвістики (рубрика)

ЛІНГВОМЕТОДОЛОГІЯ, ЛІНГВІСТИКА ТЕКСТУ

2. Реєстрове слово (слово-термін, словосполучення-термін)

ІНТРАЛІНГВІСТИКА, КОНЦЕПТУАЛЬНИЙ АНАЛІЗ

3. Дефініція – тлумачення змісту слів і словосполучень. У тезаурусі використано чотири різновиди дефініцій. Для більшості дефініцій подано також ілюстративний матеріал. Коротке визначення слова з виділенням найбільш суттєвих ознак:

АКТУАЛІЗАЦІЯ – реалізація у мовленні потенційних можливостей мовних одиниць, їх адаптація до конкретної мовленнєвої ситуації.

3.1. Дефініції, які акцентують на змістовній тотожності реєстрового слова з іншими. У визначенні подано посилання на термін-синонім: **ВІДПРАВНИК**: те, що і **АДРЕСАНТ**

3.2. Дефініції-посилання, які не пояснюють зміст терміна-слова чи терміна-словосполучення, а відсилають користувача до інших словникових статей, які містять необхідну інформацію: **КАТЕГОРІЯ ЧАСУ**: див. **ЧАС ДІЄСЛОВА**

3.3. Дефініції, які містять вказівку на те, що реєстрове слово перебуває у логічних зв'язках з іншим реєстровим словом:

КВАНТИТАТИВНА ЛІНГВІСТИКА: Розділ прикладної лінгвістики, який застосовує кількісні (статистичні) методи для створення структурно-ймовірнісної моделі мови.

КОРЕЛЯТИ: Члени кореляції, пари чи ряди взаємозумовлених мовних явищ: читати/прочитати – кореляти за видом (доконаний/недоконаний)

СЛОВОСКЛАДАННЯ – спосіб утворення нових слів шляхом з'єднання цілих слів: магазин-салон, хіт-парад, балетмейстер.

4. Термін-синонім (синоніми): **ІНТЕГРАЛЬНІСТЬ ТЕКСТУ**: (**ЦІЛІСНІСТЬ ТЕКСТУ**), **КОНТРАСТИВНИЙ АНАЛІЗ**: (**ЗІСТАВНИЙ АНАЛІЗ**)

5. Термін-корелят (кореляти): **КОГЕРЕНТНІСТЬ**, **ІНТЕРДИСКУРСИВНІСТЬ**, **ІНТЕРТЕКСТУАЛЬНІСТЬ**: (**ЦІЛІСНІСТЬ**);

ГЛИБИННА СТРУКТУРА ТЕКСТУ: (**ПОВЕРХНЕВА СТРУКТУРА**)

6. Термін-голомім: **МЕТОД**, **АНАЛІЗ** (**НАРАТИВНИЙ АНАЛІЗ**)

7. Термін-меронім: **БЕЗОСОБОВИЙ ПРИСУДОК**: (**БЕЗОСОБОВІ РЕЧЕННЯ**); **ЗМЕНШУВАЛЬНІ АФІКСИ**: (**ДИМІНУТИВ**)

8. Термін-гіперонім: **МЕТОНІМІЯ**: (**АНТОНОМАЗІЯ**); **ЛЕКСИЧНА ОДИНИЦЯ**: (**ДИСФЕМІЗМ**)

9. Термін-гіпонім: **ЛІТЕРА**: (**ГРАФЕМА**); **ПРЯМИЙ КОМУНІКАТИВНИЙ АКТ**: (**КОМУНІКАТИВНИЙ АКТ**)

Подані тут типи зв'язків між термінами лінгвістичної науки дають доволі повну картину про терміносистему лінгвістичної науки. У табл. 1 подано узагальнений вигляд кількох словникових статей з різних галузей мовознавства.

Як видно, для більшості термінів заповнені не всі рядки таблиці. Ці незаповнені місця можуть пояснюватися двоюко: немає терміна, який відповідає функції, проте не можна стверджувати, що його не існує. За визначенням С. Є. Нікітіної, це “порожні” місця у “Таблиці Менделєєва” лінгвістичних термінів [7]. Місце не заповнене, тому що для реєстрового слова така функція не істотна, хоча ймовірно, що вона може існувати для широкого родового поняття, яке перебуває від даного трохи далі у ієрархічній структурі.

Необхідно також розрізнити два підходи до тезауруса. Перший трактує тезаурус як спосіб відображення термінологічної системи лінгвістики. При другому підході тезаурус можна розглядати як інструмент індексування та інформаційного пошуку. З цією метою словникову статтю можна модифікувати до двох пунктів, враховуючи різний ступінь близькості за змістом між терміном (реєстровим) словом та його “сусідами” [9–10]. Ми пропонуємо звести усі функції до стандартного набору (табл. 2).

Таблиця 1. Узагальнений вигляд кількох словникових статей з різних галузей мовознавства

Рубрика	Термін	Дефініція. Приклад	Синонім	Корелят	Голонім	Меронім	Гіперонім	Гіпонім
ЛІНГВІСТИКА ТЕКСТУ	СКЛАДНЕ СИНТАК- СИЧНЕ ЦІЛЕ	Фрагмент тексту, який складається з кількох пов'язаних між собою речень. Засобами зв'язку компонентів є їх змістова єдність, інтонація, порядок слів, вживання особових і вказівних займенників і т.д.			Зв'язний текст	Речення		
ЛІНГВО- СЕМІОТИКА	ІДЮСТИЛЬ	Стиль художнього твору того чи іншого письменника чи всієї його творчості	Індивідуальний стиль	Функціональний стиль	Ідіолект		Стиль	
ЛІНГВО- СЕМІОТИКА	ГРАДАЦІЯ	Стилістичний прийом розташування слів і виразів за наростанням їх значущості і експресивності: <i>"Ти ж повертайся, вертайся додому... додому... Ждана повернення вибори, вимолі, виболій..."</i>					Стилістичний прийом	Клімакс, антиклімакс
ЛІНГВОМЕТО- ДОЛОГІЯ	АСОЦІАТИВ- НИЙ ЕКСПЕ- РИМЕНТ	Один із лінгвістичних методів, який полягає у визначенні мовних реакцій на слова-стимули та зв'язків між відчуттями, почуттями та їх відображень у ментальному лексиконі мовців. Передбачає анкетування інформантів зі списками слів-стимулів і визначення найчастотніших реакцій			Методика	Анкетування респондентів	Лінгвістична процедура, метод	
ЛІНГВОМЕТО- ДОЛОГІЯ	ПОРІВНЯЛЬ- НО-ІСТОРИЧ- НЕ МОВО- ЗНАВСТВО	Розділ мовознавства, який займається порівняльно-історичним вивченням генетично споріднених мов для визначення їх схожості чи відмінності на всіх рівнях	Компаративістика		Мовознавство, лінгвістика		Наукова дисципліна	
ЛЕКСИКО- ЛОГІЯ	ТОПО- НІМІКА	Розділ ономастики, який вивчає закономірності виникнення, розвитку та функціонування топонімів	Психолінгвістичний експеримент		Лексикологія		Ономастика	Мікротопоніміка, Макротопоніміка, Гідроніміка, Ойконіміка, Ороніміка, Урбоніміка

**Таблиця 2. Набір лінгвістичних відношень
(мероніми, гіпоніми, синоніми, гіпероніми, голоніми)**

Ступінь близькості	Назва відношення
“Близько”	Вид, компонент (мероніми, гіпоніми), синоніми
“Далеко”	Кореляти, рід (гіпероніми), ціле (голоніми)

Розроблений таким чином тезаурус є базою для побудови онтології предметної області прикладної лінгвістики, оскільки в ньому визначені основні терміни та зв'язки між ними.

ВИСНОВКИ

Описану модель тезауруса і методику його побудови розглянуто у двох аспектах: як лінгвістичний опис інформаційного пошукового тезауруса і як інформаційно-логічний опис лінгвістичної термінології. Опис методики створення тезауруса містить матеріал, який можна використати для часткової формалізації побудови словникової статті. Якщо існує перелік зв'язків, які вносять у тезаурус, є перелік визначень, які описують ці зв'язки, словник, одиниці якого є кандидатами в терміни, і є тексти, у яких визначено понятійні зв'язки між термінами, то можливим є, дотримуючись інструкції (яка співвідносить текстовий опис зв'язків з переліком цих зв'язків), заповнити для кожного терміна його словникову статтю.

У поданому тезаурусі намагалися знайти компроміс між кількістю зв'язків, необхідних для відображення семантичної системи лінгвістичної термінології і водночас прийнятних за кількістю для інформаційних цілей. Проведене дослідження є одним із перших кроків на шляху побудови онтології предметної області прикладної лінгвістики, а також вивчення понятійних зв'язків лінгвістичної термінології прикладними методами, а тезаурус містить матеріал як для подальших досліджень мови науки, так і для лінгвістичної науки зокрема.

1. *Караулов Ю. Н.* Лингвистическое конструирование и тезаурус литературного языка. – М.: Наука, 1981. – 367 с.
2. *Заседателева М. Г.* Тезаурусное моделирование предметной области “компетентностный подход”: дизайн исследования и программная реализация на платформе ИС “Семограф” // Инновационные проекты и программы в образовании. – 2013. – № 4. – С. 3–11.
3. *Земляная Т. Б., Павлычева О. Н.* Терминологические словари и тезаурусы // Журн. научно-педагогической информации. – 2010. – № 4.
[Электронный ресурс] / <http://www.paedagogia.ru/2010/41-04/104-zemlyanayapavlicheva>.
4. *Ковязина М. А.* Тезаурусное описание экологической терминосистемы // Единство системного и функционального анализа языковых единиц: материалы Междунар. научн. конф.: в 2 ч. – Белгород: Изд-во БелГУ, 2006. – Вып. 9. – Ч. II. – С. 358–362.
5. *Сергеева Т. Н.* Тезаурусное моделирование предметной области “корпусная лингвистика” на материале английского языка // Вектор науки ТГУ. – 2009. – № 1(4). – С. 87–90.
6. *Горбунов Е. Ю.* Тезаурусный подход к моделированию английской морфологической терминологии // Вестник СамГУ. – 2006. – № 10/2 86 (50). – С. 82–89.
7. *Никитина С. Е.* Тезаурус по теоретической и прикладной лингвистике (Автоматическая обработка текста). – М.: Наука, 1978. – 374 с.
8. *Селіванова О. О.* Сучасна лінгвістика: напрями та проблеми. – Полтава: Довкілля, 2008. – 712 с.
9. *Мороз О. В.* Особливості бібліометричних досліджень у гуманітарних науках // Зб. наук. праць. Людина. Комп'ютер. Комунікація. – Львів: НУ “Львівська політехніка”, 2010. – С. 173–174.
10. *Мороз О. В.* Сучасний стан бібліометричних досліджень: зарубіжний та вітчизняний досвід // Мова і культура. – К.: Видавничий дім Дмитра Бураго, 2011. – Вип. 14. – Т. (III) (148). – С. 212–217.